Multivariate Optimization

David Carlson

2021

Multivariate Optimization

Overview

- Think back to when we discussed first and second derivatives of a one-dimensional funciton.
- Combined, the first and second derivatives can tell us whether a point is a maximum or minimum of $f(x)$.

$f'(x) = 0$ and $f''(x) < 0$ $f'(x) = 0$ and $f''(x) > 0$ $f'(x) = 0$ and $f''(x) = 0$

Conditions for Extrema in $\mathbf{R}^n$

- The conditions for extrema are similar to those for functions on $\mathbf{R}^1$.
- Let $f(\mathbf{x})$ be a function of n variables.
- Let $B(\mathbf{x}, \epsilon)$ be the ϵ-ball about the point $\mathbf{x}$. Then
  1. $f(\mathbf{x}^*) > f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$
     $\implies$ Strict Local Max
  2. $f(\mathbf{x}^*) \geq f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$
     $\implies$ Local Max
  3. $f(\mathbf{x}^*) < f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$
     $\implies$ Strict Local Min
  4. $f(\mathbf{x}^*) \leq f(\mathbf{x}),\ \forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$
     $\implies$ Local Min

First order conditions

- When we examined functions of one variable x, we found critical points by taking the first derivative, setting it to zero, and solving for x.
- For functions of n variables, the critical points are found in much the same way, except now we set the partial derivatives equal to zero.
- Note: We will only consider critical points on the interior of a function's domain.
- $\mathbf{x}^*$ is a critical point iff $\nabla f(\mathbf{x}^*) = 0$.

Example

Find the critical points of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$

1. The partial derivatives of $f(\mathbf{x})$ are

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial x_1} &= 2(x_1 - 1) \\ \frac{\partial f(\mathbf{x})}{\partial x_2} &= 2x_2\end{aligned}$$

2. Setting each partial equal to zero and solving for x_1 and x_2, we find that there's a critical point at $\mathbf{x}^* = (1, 0)$.

Second order conditions

- When we found a critical point for a function of one variable, we used the second derivative as an indicator of the curvature at the point in order to determine whether the point was a min, max, or saddle.
- For functions of n variables, we use second order partial derivatives as an indicator of curvature.

Curvature and The Taylor Polynomial as a Quadratic Form

- The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.

Curvature and The Taylor Polynomial as a Quadratic Form

- The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.
- The second order **Taylor polynomial** about the critical point $\mathbf{x}^*$ is

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T \mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

Curvature and The Taylor Polynomial as a Quadratic Form

- The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.
- The second order **Taylor polynomial** about the critical point $\mathbf{x}^*$ is

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

- Since we're looking at a critical point, $\nabla f(\mathbf{x}^*) = 0$; and for small $\mathbf{h}$, $R(\mathbf{h})$ is negligible. Rearranging, we get

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

Curvature and The Taylor Polynomial as a Quadratic Form

- The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at $\mathbf{x}$ — e.g., which tells us whether a critical point $\mathbf{x}^*$ is a min, max, or saddle point.
- The second order **Taylor polynomial** about the critical point $\mathbf{x}^*$ is

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

- Since we're looking at a critical point, $\nabla f(\mathbf{x}^*) = 0$; and for small $\mathbf{h}$, $R(\mathbf{h})$ is negligible. Rearranging, we get

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

- The RHS is a quadratic form and we can determine the definiteness of $\mathbf{H}(\mathbf{x}^*)$.

Breaking that down

- We want to know if $f(\mathbf{x})$ is increasing or decreasing around point $\mathbf{x}^*$.

Breaking that down

- We want to know if $f(\mathbf{x})$ is increasing or decreasing around point $\mathbf{x}^*$.
- We want to know if it is increasing or decreasing in any direction. So we can calculate:

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*)$$

for a generic $\mathbf{h}$.

Breaking that down

- We want to know if $f(\mathbf{x})$ is increasing or decreasing around point $\mathbf{x}^*$.
- We want to know if it is increasing or decreasing in any direction. So we can calculate:

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*)$$

  for a generic $\mathbf{h}$.
- Is this going to be always positive, always negative, or will it depend on the specific value of $\mathbf{h}$?

- To figure this out, we are going to rely on something called a Taylor series approximation.
- In one dimension, it turns out that you can approximate any function evaluated *near* (but not at) point a as:

$$f(x) \approx \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$$

- Example:

$$f(x) \approx f(0) + \frac{f'(0)}{1!}(x-0) + \frac{f''(0)}{2!}(x-0)^2 + \frac{f'''(0)}{3!}(x-0)^3 \dots$$

- Further, since $(x-a)$ is going to be small, the higher order polynomials such as $(x-a)^4$ start heading to zero. So, we can often just write this as:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + R(x)$$

- Or even:

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2$$

- This is called a second order Taylor polynomial

Second order Taylor polyomial and Hessians

- In our problem use a second order taylor polynomial for higher dimensions:

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^T \mathbf{H}(\mathbf{x}^*)\mathbf{h} + R(\mathbf{h})$$

- We know already that the second term is zero (that's how we identified the critical point)
- We now re-organize

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^T \mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

Definiteness and Hessians

- So we need to characterize the following function:

$$\frac{1}{2}\mathbf{h}^T\mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

  - $\mathbf{H}$ is a matrix
  - $\mathbf{h}$ is a vector

Definiteness

- When some $n \times n$ matrix $\mathbf{A}$ is pre- and post-multiplied by a conformable non-zero matrix $\mathbf{x}$, we get the equation:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = c$$

- In one dimension, this would be:

$$c = xax$$

$$c = ax^2$$

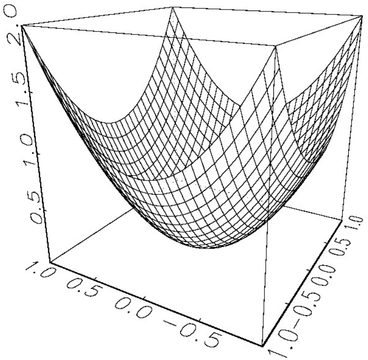
- In more than one dimension it would be a quadratic formula

Defining definite

- For all nonzero vectors **x**:
  1. **A** is said to be **positive definite** if $c > 0$.
  2. **A** is said to be **positive semidefinite** if $c \geq 0$.
  3. **A** is said to be **negative definite** if $c < 0$.
  4. **A** is said to be **negative semidefinite** if $c \leq 0$.
  5. **A** is **indefinite** if none of these apply.

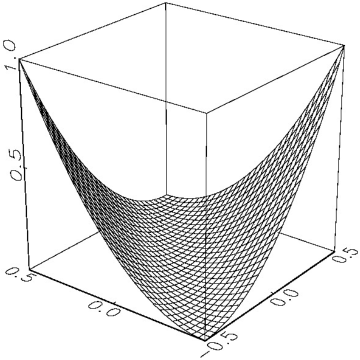
Examples: Positive Definite

- $$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} \\ &= x_1^2 + x_2^2 \end{aligned}$$



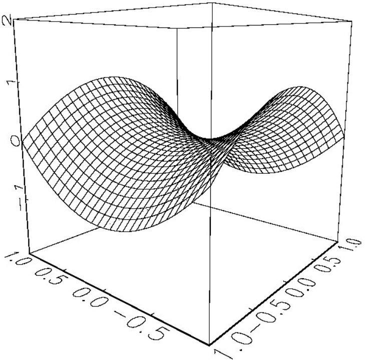
Example: Positive Semidefinite

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{x}$$
$$= (x_1 - x_2)^2$$



Example: Indefinite

$$
\begin{aligned}
Q(\mathbf{x}) &= \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{x} \\
&= x_1^2 - x_2^2
\end{aligned}
$$



Tests for definiteness

- Now we have defined it, how do we prove it?
- How can we test if a specific matrix meets one of these criteria?
- We've got one, but first we need to define terms.

Defining terms: Principal minors

- Given an $n \times n$ matrix $\mathbf{A}$, kth order **principal minors** are the determinants of the $k \times k$ submatrices along the diagonal obtained by deleting $n - k$ columns and the same $n - k$ rows from $\mathbf{A}$.

- Example: For a 3×3 matrix $\mathbf{A}$,
  1. First order principal minors:
  $$|a_{11}|, \quad |a_{22}|, \quad |a_{33}|$$
  2. Second order principal minors:
  $$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$
  3. Third order principal minor: $|\mathbf{A}|$

Defining terms: Leading principal minors

- Define the kth **leading principal minor** M_k as the determinant of the $k \times k$ submatrix obtained by deleting the *last* $n - k$ rows and columns from $\mathbf{A}$.
- Example: For a 3×3 matrix $\mathbf{A}$, the three leading principal minors are

$$M_1 = |a_{11}|, \quad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

An algorithm for determining definiteness of a matrix

If **A** is an $n \times n$ symmetric matrix, then

1. $M_k > 0,\ k = 1, \dots, n \quad \Longrightarrow \quad$ Positive Definite
2. $M_k < 0$, for odd k and $M_k > 0$, for even $k \quad \Longrightarrow$ Negative Definite
3. $M_k \neq 0,\ k = 1, \dots, n$, but does not fit the pattern of 1 or 2. $\Longrightarrow \quad$ Indefinite.

Finding semidefinite matrices

If some leading minor is equal to zero, but the others fit the patterns in 1 or 2 above:

1. Every principal minor $\geq 0 \qquad \Longrightarrow \qquad$ Positive Semidefinite
2. Every principal minor of odd order ≤ 0 and every principal minor of even order $\geq 0 \qquad \Longrightarrow \qquad$ Negative Semidefinite

Returning to optimization

- To determine whether a critical point is a global min or max, we can check the concavity of the function over its entire domain.
- Here again we use the definiteness of the Hessian to determine whether a function is globally concave or convex:
  1. $\mathbf{H}(\mathbf{x})$ Positive Semidefinite $\forall \mathbf{x} \implies$ Globally Convex
  2. $\mathbf{H}(\mathbf{x})$ Negative Semidefinite $\forall \mathbf{x} \implies$ Globally Concave
- Notice that the definiteness conditions must be satisfied over the entire domain.

- Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$,
  1. $\implies$ Global Min
  2. $\implies$ Global Max

Global, local, and semidefinite Hessians

- Note that showing that $\mathbf{H}(\mathbf{x}^*)$ is negative semidefinite is not enough to guarantee $\mathbf{x}^*$ is a global max.
- However, showing that $\mathbf{H}(\mathbf{x})$ is negative semidefinite for all $\mathbf{x}$ guarantees that x^* is a global max.
- (The same goes for positive semidefinite and minima.)

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.
- Both have $x = 0$ as a critical point.

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.
- Both have $x = 0$ as a critical point.
- Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either.

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.
- Both have $x = 0$ as a critical point.
- Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either.
- However, $f_1''(x) = 12x^2$ and $f_2''(x) = -12x^2$.
- For all x, $f_1''(x) \geq 0$

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.
- Both have $x = 0$ as a critical point.
- Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either.
- However, $f_1''(x) = 12x^2$ and $f_2''(x) = -12x^2$.
- For all x, $f_1''(x) \geq 0$ and $f_2''(x) \leq 0$ — i.e., $f_1(x)$ is globally convex and $f_2(x)$ is globally concave.

- Take $f_1(x) = x^4$ and $f_2(x) = -x^4$.
- Both have $x = 0$ as a critical point.
- Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either.
- However, $f_1''(x) = 12x^2$ and $f_2''(x) = -12x^2$.
- For all x, $f_1''(x) \geq 0$ and $f_2''(x) \leq 0$ — i.e., $f_1(x)$ is globally convex and $f_2(x)$ is globally concave.
- So $x = 0$ is a global min of $f_1(x)$ and a global max of $f_2(x)$.

Example

Given $f(\mathbf{x}) = x_1^3 - x_2^3 + 9x_1x_2$, find any maxima or minima.

1. First-order conditions. Set the gradient equal to zero and solve for x_1 and x_2.

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= 3x_1^2 + 9x_2 = 0 \\
\frac{\partial f}{\partial x_2} &= -3x_2^2 + 9x_1 = 0
\end{aligned}$$

Example

Given $f(\mathbf{x}) = x_1^3 - x_2^3 + 9x_1x_2$, find any maxima or minima.

1. First-order conditions. Set the gradient equal to zero and solve for x_1 and x_2.

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= 3x_1^2 + 9x_2 = 0 \\
\frac{\partial f}{\partial x_2} &= -3x_2^2 + 9x_1 = 0
\end{aligned}$$

   We have two equations in two unknowns.

Solving for x_1 and x_2, we get two critical points: $\mathbf{x}_1^* = (0, 0)$ and $\mathbf{x}_2^* = (3, -3)$.

2. Second order conditions. Determine whether the Hessian is positive or negative definite.

The Hessian is

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 6x_1 & 9 \\ 9 & -6x_2 \end{pmatrix}$$

2. Second order conditions. Determine whether the Hessian is positive or negative definite.

The Hessian is

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 6x_1 & 9 \\ 9 & -6x_2 \end{pmatrix}$$

Evaluated at $\mathbf{x}_1^*$,

$$\mathbf{H}(\mathbf{x}_1^*) = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}$$

The two leading principal minors are $M_1 = 0$ and $M_2 = -81$, so $\mathbf{H}(\mathbf{x}_1^*)$ is indefinite and $\mathbf{x}_1^* = (0, 0)$ is a saddle point.

Evaluated at $\mathbf{x}_2^*$,

$$\mathbf{H}(\mathbf{x_2^*}) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

The two leading principal minors are $M_1 = 18$ and $M_2 = 243$.

Evaluated at $\mathbf{x}_2^*$,

$$\mathbf{H}(\mathbf{x_2^*}) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

The two leading principal minors are $M_1 = 18$ and $M_2 = 243$. Since both are positive, $\mathbf{H}(\mathbf{x_2^*})$ is positive definite and $\mathbf{x}_2^* = (3, -3)$ is a strict local min.

Find and characterize the extrema and sadle points for the the following functions.

1. $$f(x,y) = \frac{3}{2}x^2 - 2xy - 5x + 2y^2 - 2y$$

2. $$f(x,y,z) = -3x^2 - 2xy + xz - \frac{1}{2}y^2 - yz - 4z^2 + 5x + 7y + 25z$$

3. $$f(x,y) = x^2 + 6xy + y^2 - 18x - 22y + 5$$